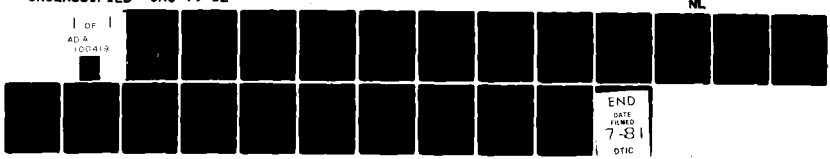
AD-A100 419 CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER F/G 12/1
STOCHASTICALLY-ORDERED PARAMETERS IN BAYESIAN PREDICTION.(U)
OCT 79 W S JEWELL AFOSR-81-0122
UNCLASSIFIED ORC-79-12 NL



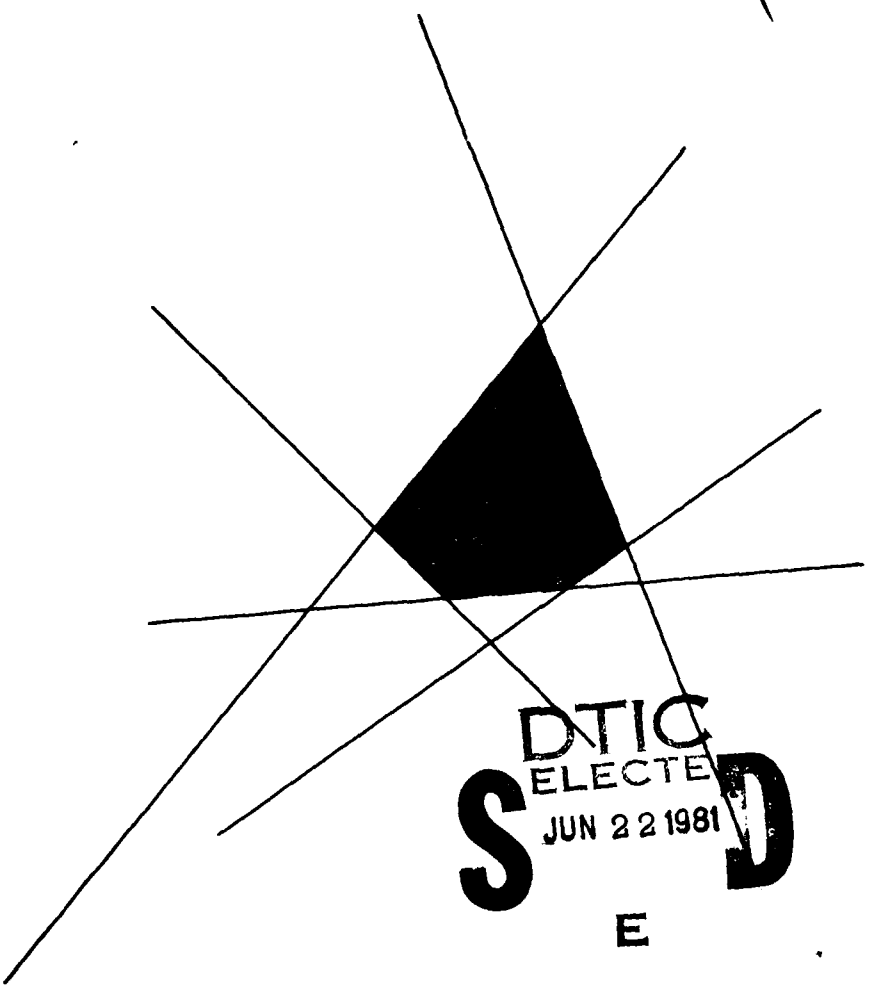
ORC 79-12
OCTOBER 1979

LEVEL II

12

STOCHASTICALLY-ORDERED PARAMETERS IN BAYESIAN PREDICTION

by
WILLIAM S. JEWELL



OPERATIONS
RESEARCH
CENTER

UNIVERSITY OF CALIFORNIA · BERKELEY

STOCHASTICALLY-ORDERED PARAMETERS IN BAYESIAN PREDICTION

by

William S. Jewell
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

OCTOBER 1979 ORC 79-12

This research has been partially supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grants AFOSR-81-0122 and AFOSR-77-3179 with the University of California.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC 79-12	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) STOCHASTICALLY-ORDERED PARAMETERS IN BAYESIAN PREDICTION		5. TYPE OF REPORT & PERIOD COVERED Research Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William S. Jewell		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0122, [illegible]
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Office of Scientific Research Bolling Air Force Base, D.C. 20332		12. REPORT DATE October 1979 13. NUMBER OF PAGES 21
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-77-3179.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Reliability Growth
Stochastic Ordering
Bayesian Prediction

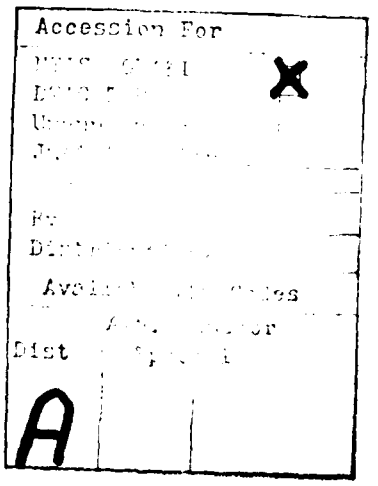
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)



ABSTRACT

In models of reliability growth in stages, it is usual to assume that system parameters improve monotonically from stage to stage, following some postulated law of growth. This paper explores a Bayesian model where such improvement only occurs on the average, e.g., a case when the parameters are assumed to be *stochastically* ordered. It is shown that the problem can be recast into a hierarchical form in which there are strictly-ordered hyperparameters which index the admissible family of ordered distributions for the parameters; the modelling problem is then to describe an appropriate law of motion over the hyperparameters.

STOCHASTICALLY-ORDERED PARAMETERS IN BAYESIAN PREDICTION

by

William S. Jewell

0. INTRODUCTION

Models of *reliability growth* have received increasing attention in recent literature. One approach is to assume an underlying continuous improvement in system parameters; we call this the *learning-curve* approach (Jewell [2]). A second model, which may be more appropriate under certain circumstances, is *growth by stages*, in which the parameters increase (or decrease) monotonically at certain fixed points in time, when design improvements are made. In both of these models, an important problem is the *prediction of the ultimate performance* of the system after all reliability growth has occured or all the design modifications have been effected; in practical procurement problems, it is desirable to make these predictions using as little test data as possible.

This paper explores an extension of the growth by stages model suggested by N. Singpurwalla [3] in which the parameters improve only *stochastically*, rather than absolutely.

I. DETERMINISTICALLY-ORDERED PARAMETERS

Consider the usual set-up, in which *random observables* $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_t$ in periods $1,2, \ldots, t$ are governed by *random parameters* according to some *likelihood density* $p(x_1, x_2, \ldots, x_t \mid \theta_1, \theta_2, \ldots, \theta_t)$. Given a *prior parameter density* $p(\theta_1, \theta_2, \ldots, \theta_t)$, and *observations* $\underline{x} = (x_1, x_2, \ldots, x_t)$, a straightforward application of Bayes' law calculates the *posterior parameter density* $p(\theta_1, \theta_2, \ldots, \theta_t \mid \underline{x})$.

In most applications, $\tilde{\theta}_t$ is assumed sufficient for $\tilde{x}_t$, so that the likelihood takes the simpler form:

$$(1.1) \qquad p(x_1, x_2, \ldots, x_t \mid \theta_1, \theta_2, \ldots, \theta_t) = \prod_{i=1}^{t} p_i(x_i \mid \theta_i) \; ;$$

we retain explicit dependence upon the time period primarily to permit stages of different duration, but not usually different failure laws in each interval.

In the usual growth-by-stages model, the improvement is expressed in terms of a deterministic ordering of the parameters, say:

$$(1.2) \qquad \theta_1 \geq \theta_2 \geq \cdots \geq \theta_t \; ,$$

and this is handled in an obvious way through the definition of the parameter prior and its support. A usual simplification is to fix some θ_0, and then suppose that $\tilde{\theta}_{i+1}$ depends only upon $\tilde{\theta}_i$, and not upon other parameters or the index i, so that:

$$(1.3) \quad p(\theta_1, \theta_2, \ldots, \theta_t) = \prod_{i=1}^{t} p(\theta_i \mid \theta_{i-1}) \qquad (\theta_0 \geq \theta_1 \geq \theta_2 \geq \theta_t) \; .$$

Typically a prototype parameter density $g(u)$, $0 \leq u \leq 1$, is selected for modelling convenience, giving a common form to the "shrinkage" of the parameters, viz.

$$(1.4) \qquad p(\theta_i \mid \theta_{i-1}) = \frac{1}{\theta_{i-1}} g\left(\frac{\theta_i}{\theta_{i-1}}\right) \qquad (0 \leq \theta_i \leq \theta_{i-1})$$

$$(i = 1,2, \ldots, t) .$$

The actual Bayesian calculations are complicated, but present no conceptual difficulties; Smith [4] gives an example with $p(x_t \mid \theta_t)$ binomial, and $g(\cdot)$ uniform. Of course, (1.2), (1.3) and (1.4) represent rather strong beliefs about the underlying growth mechanism.

2. STOCHASTICALLY-ORDERED PARAMETERS

N. Singpurwalla [3] has suggested that (1.2) might reasonably be replaced by the weaker hypothesis that

$$\tilde{\theta}_1 \gtrsim \tilde{\theta}_2 \gtrsim \cdots \gtrsim \tilde{\theta}_t \ . \tag{2.1}$$

This might mean, for example, that while the failure rate *tended* to decrease from stage to stage, such growth was by no means certain in a given realization. We now modify the basic model to handle this assumption.

(2.1) implies that the parameters are selected from some family F of priors which are *stochastically ordered* over the different stages. Let α be a *scalar hyperparameter* which indexes the members of this family. Then we can associate $\alpha_1, \alpha_2, \ldots, \alpha_t$ with stages $1,2, \ldots, t$, and arrange that

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_t \tag{2.2}$$

(or an increasing version) selects an appropriately ordered distribution for each parameter. Specifically, F is a family of *conditional prior densities* $p(\theta \mid \alpha)$, with complementary distribution $P^c(\theta \mid \alpha)$, such that, for every value of θ and every $i < j$ $(i, j = 1,2, \ldots)$

$$P^c(\theta \mid \alpha_i) \geq P^c(\theta \mid \alpha_j) \ ; \tag{2.3}$$

this can be visualized as in Figure 1. α_t can be an actual hyperparameter of the (conditional) prior, or it can be an abstract index that selects one of an admissible member of ordered priors. The key point is to arrange things so that (2.2) guarantees (2.1).

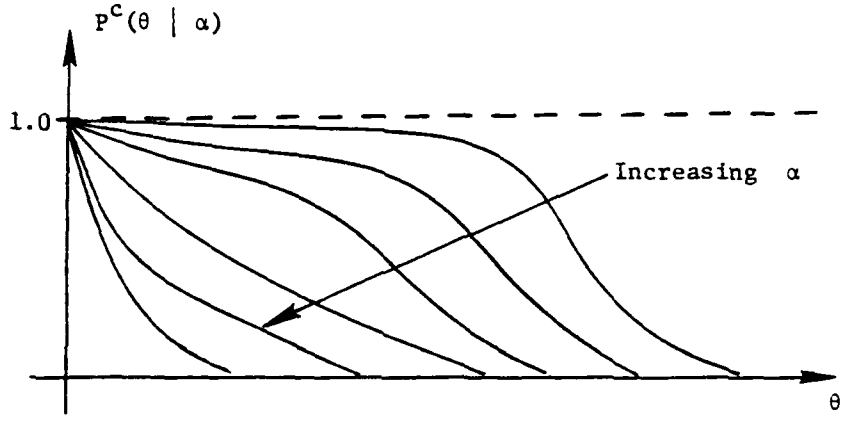


FIGURE 1

ADMISSIBLE FAMILY, F , OF STOCHASTICALLY

ORDERED CONDITIONAL PRIORS ON θ

With (2.1) guaranteed, what can we say about the selection of the hyperparameters? In a Bayesian model we must now express our prior beliefs and previous experience about what values of $\alpha_1, \alpha_2, \ldots, \alpha_t$ might occur during the different stages. In other words, we consider the hyperparameters as random variables with a given *hyperprior density* $p(\alpha_1, \alpha_2, \ldots, \alpha_t)$ over the family F satisfying (2.2).

Effectively, the stochastic ordering has introduced a *hierarchical Bayesian formulation* (see, e.g., Jewell, [1]), in which we again have deterministic ordering, this time in the hyperparameters. Note also that this approach avoids the "pseudo-Bayesian" formulation described by Singpurwalla [3].

For convenience in the sequel, we assume a Markovian hyperprior similar to (1.3), so that, given α_0, and a family of (hyperprior) *transition densities* $p(\alpha_i \mid \alpha_{i-1})$, we have:

$$(2.4)\quad p(\alpha_1, \alpha_2, \ldots, \alpha_t) = \prod_{i=1}^{t} p(\alpha_i \mid \alpha_{i-1}) \qquad (\alpha_0 \geq \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_t) .$$

(α_0 could, of course, have an initial density of its own. But, for simplicity, we omit further explicit mention of α_0).

In summary, we have the following model:

(a) Prior beliefs or experience specify:

(i) α_0 and the hyperprior transition densities $p(\alpha_i \mid \alpha_{i-1})$ with deterministic ordering (2.2);

(ii) the family F of admissible conditional priors, so that, given α_i, $\tilde{\theta}_i$ has complementary distribution $P^c(\theta_i \mid \alpha_i)$ and stochastic ordering (2.1);

(b) A known failure law governs the different observational likelihoods $p_i(x_i \mid \theta_i)$.

Note that $\tilde{\alpha}_i$ is sufficient for $\tilde{\theta}_i$ which is, in turn sufficient for $\tilde{x}_i$, so that, given $\underline{\alpha} = (\alpha_1, \ldots, \alpha_t)$, the problem of predicting $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_t)$ and/or $\underline{x} = (x_1, x_2, \ldots, x_t)$ decomposes into several separate problems. However, the usual situation is that we are *given* the observations $\underline{x}$, and want to *update* the distributions of the unobserved $\underline{\tilde{\theta}}$ and $\underline{\tilde{\alpha}}$, or to *predict* future observations $\tilde{x}_{t+1}, \tilde{x}_{t+2}, \ldots$. Because of the Markovian dependence of the $\tilde{\alpha}_i$, both of these problems require extensive applications of Bayes' law.

3. MARGINAL DENSITIES

As a preliminary, we indicate how to calculate the various marginal densities. For the parameters, the joint marginal is:

$$(3.1)\quad p(\theta_1,\theta_2,\ \ldots,\ \theta_t) = \iint\cdots\int \prod_{i=1}^{t} p(\theta_i \mid \alpha_i)p(\alpha_i \mid \alpha_{i-1})d\alpha_i \ .$$

Perhaps the easiest way to calculate is by defining:

$$(3.2)\qquad p(\theta_1;\alpha_2) = \int p(\theta_1 \mid \alpha_1)p(\alpha_2 \mid \alpha_1)p(\alpha_1 \mid \alpha_0)d\alpha_1 \ ,$$

and then proceed iteratively to find:

$$(3.3)\qquad \begin{aligned} p(\theta_1,\theta_2,\ \ldots,\ \theta_i\ ;\ \alpha_{i+1}) &= \int p(\theta_i \mid \alpha_i)p(\alpha_{i+1} \mid \alpha_i) \\ &\quad \cdot\ p(\theta_1,\theta_2,\ \ldots,\ \theta_{i-1}\ ;\ \alpha_i)d\alpha_i \end{aligned}$$

for $i = 2,3,\ \ldots$. At any point t where the marginal density is desired,

$$(3.4)\quad p(\theta_1,\theta_2,\ \ldots,\ \theta_t) = \int p(\theta_t \mid \alpha_t)\ \cdot\ p(\theta_1,\theta_2,\ \ldots,\ \theta_{t-1}\ ;\ \alpha_t)d\alpha_t \ .$$

It is clear that the θ_i's constitute a stochastic process which is *not* Markovian (as in (1.3)). One can show, however, that the nature of F, expressed in (2.3), implies ordered marginals:

$$(3.5)\qquad \Pr\{\tilde{\theta}_i \geq \theta\} \geq \Pr\{\tilde{\theta}_j \geq \theta\}\ ,\ (i < j)\quad (\forall\ \theta)$$

as expected.

For the observables $\{\tilde{x}_t\}$, the intermediate nature of the $\{\tilde{\theta}_i\}$ suggests we define *"bypass" conditional densities* from the likelihood and prior, viz.,

$$p_i(x_i \mid \alpha_i) = \int p_i(x_i \mid \theta_i)p(\theta_i \mid \alpha_i)d\theta_i \tag{3.6}$$

for $i = 1,2, \ldots$. The joint marginal of the observations is then similar to (3.1):

$$p(x_1,x_2, \ldots, x_t) = \iint \cdots \int \prod_{i=1}^{t} p_i(x_i \mid \alpha_i)p(\alpha_i \mid \alpha_{i-1})d\alpha_i \ , \tag{3.7}$$

which can also be "bootstrapped" as in (3.2), (3.3). Alternately, if (3.1) or (3.4) were first calculated,

$$p(x_1,x_2, \ldots, x_t) = \iint \cdots \int p(\theta_1,\theta_2, \ldots, \theta_t) \prod_{i=1}^{t} p_i(x_i \mid \theta_i)d\theta_i \ . \tag{3.8}$$

4. UPDATING HYPERPARAMETERS

We turn now to the calculation of the *inverse probabilities*. By using the bypass conditional densities of (3.6), we find that (given α_0):

$$(4.1)\quad p(\alpha_1, \ldots, \alpha_t \,;\, x_1, x_2, \ldots, x_t) = \prod_{i=1}^{t} p_i(x_i \mid \alpha_i) \cdot p(\alpha_i \mid \alpha_{i-1})$$

so that the *update* formula for the hyperparameters (i.e., the "hyper-posterior-to-the-data" density) is:

$$(4.2)\qquad p(\alpha_1, \ldots, \alpha_t \mid \underline{x}) = \frac{\prod_{i=1}^{t} p_i(x_i \mid \alpha_i) \cdot p(\alpha_i \mid \alpha_{i-1})}{p(\underline{x})} .$$

$p(\underline{x})$ is calculated in (3.7), but in most practical calculations, it is treated simply as a normalization factor. Note that "later" observables provide information about "earlier" hyperparameters because of the Markovian dependence.

5. PREDICTING HYPERPARAMETERS

The prediction of $\tilde{\alpha}_{t+1}$ from $\underline{x}$ is most easily "bootstrapped" in a manner similar to (3.2), (3.3). We define:

$$p(x_1;\alpha_2) = \int p_1(x_1 \mid \alpha_1)p(\alpha_2 \mid \alpha_1)p(\alpha_1 \mid \alpha_0)d\alpha_1 , \tag{5.1}$$

and then proceed iteratively for $i = 2,3, \ldots$.

$$p(x_1,x_2, \ldots, x_i ; \alpha_{i+1}) = \int p_i(x_i \mid \alpha_i)p(\alpha_{i+1} \mid \alpha_i) \cdot p(x_1,x_2, \ldots, x_{i-1} ; \alpha_i)d\alpha_i . \tag{5.2}$$

At any stage t , the prediction of $\tilde{\alpha}_t$ is a simple normalization, viz.

$$p(\alpha_{t+1} \mid \underline{x}) = \frac{p(x_1,x_2, \ldots, x_t ; \alpha_{t+1})}{p(\underline{x})} . \tag{5.3}$$

The iterative calculations (5.1), (5.2) provide also a simple formula for updating the most recent hyperparameter:

$$p(\alpha_t \mid \underline{x}) = \frac{p_t(x_t \mid \alpha_t) \cdot p(x_1,x_2, \ldots, x_{t-1} ; \alpha_t)}{p(\underline{x})} . \tag{5.4}$$

6. UPDATING PARAMETERS

It does not seem easy to provide updating for the parameters alone, since $p(\theta_1, \ldots, \theta_t \mid \underline{x})$ requires a complex integration over the Markovian transition probabilities. If *all* the marginals (3.2), (3.3) have been calculated to give (3.4), then, of course,

$$(6.1)\quad p(\theta_1,\theta_2, \ldots, \theta_t \mid \underline{x}) = \frac{\prod_{i=1}^{t} p_t(x_t \mid \theta_t)p(\theta_1,\theta_2, \ldots, \theta_t)}{p(\underline{x})} .$$

For the latest parameter only, the "bootstrap" in (5.2) can be used with

$$(6.2)\quad p(\theta_t \mid \underline{x}) = \frac{\int p_t(x_t \mid \theta_t)p(\theta_t \mid \alpha_t)p(x_1,x_2, \ldots, x_{t-1} ; \alpha_t)d\alpha_t}{p(\underline{x})} .$$

All data must be used to update $\tilde{\theta}_t$. However, if we also know α_t , then only the latest observation is sufficient, as can be seen from:

$$(6.3)\quad p(\theta_t \mid \alpha_t , \underline{x}) = \frac{p_t(x_t \mid \theta_t)p(\theta_t \mid \alpha_t)}{p(x_t \mid \alpha_t)} = p(\theta_t \mid \alpha_t , x_t) .$$

7. OTHER PREDICTIONS

In fact, the *hyperparameter forecast density* (5.3) is the key to all other predictions, since, given $\underline{x}$, only $\tilde{\alpha}_{t+1}$ is sufficient for the next period, since:

$$
\begin{aligned}
&p(x_{t+1}, \theta_{t+1}, \alpha_{t+1} \mid \underline{x}) \\
&= p_{t+1}(x_{t+1} \mid \theta_{t+1}) p(\theta_{t+1} \mid \alpha_{t+1}) p(\alpha_{t+1} \mid \underline{x}) ;
\end{aligned} \tag{7.1}
$$

marginal versions follow by integration.

For stages in the future, this generalizes to

$$
\begin{aligned}
&p(x_{t+u}, \theta_{t+u}, \alpha_{t+u} \mid \underline{x}) \\
&= p_{t+u}(x_{t+u} \mid \theta_{t+u}) p(\theta_{t+u} \mid \alpha_{t+u}) \\
&\cdot \int p^{(u-1)}(\alpha_{t+u} \mid \alpha_{t+1}) p(\alpha_{t+1} \mid \underline{x}) d\alpha_{t+1}
\end{aligned} \tag{7.2}
$$

for $u = 2,3, \ldots$, where $p^{(n)}(\cdot \mid \cdot)$ is the usual *n-step Markovian transition probability*.

8. MODEL VARIATIONS

The model described in (2.5) is actually quite general. First of all, it includes the deterministically-ordered parameter model of Section 1 by making the conditional prior, $p(\theta \mid \alpha)$, a degenerate density--i.e., $\theta_t \equiv \alpha_t$ for all t; the transition densities of (2.4) are then those of (1.3).

If, on the other hand, we assume a *time-invariant hyperparameter* $\tilde{\alpha} = \tilde{\alpha}_0 = \tilde{\alpha}_1 = \ldots = \tilde{\alpha}_t$, with initial hyperprior density $\pi(\alpha)$, then the model reduces to a *hypothesis mixing formulation*, in which data from successive trials is helping us to select one of a (usually finite) number of priors $p(\theta_t \mid \alpha)$. The parameter $\tilde{\theta}_t$ is usually also stationary, and the calculations reduce to updating *posterior to data mixing probabilities* $\pi(\alpha \mid \underline{x})$.

The strict ordering (2.2) only applies if there is a scalar hyperparameter in the prior, (or if it is an abstract index over F). Figure 2 shows the situation if the hyperparameter is a two-vector, $\underline{\alpha}_t = (\alpha_{1,t}, \alpha_{2,t})$. From the form of the conditional prior $p(\theta \mid \underline{\alpha})$, we can in principle find the *allowed region*, $R(\underline{\alpha}_{t-1})$, for $\underline{\alpha}_t$ such that $\tilde{\theta}_t \precsim \tilde{\theta}_{t-1}$, as shown. This region depends upon $\underline{\alpha}_{t-1}$ and includes it; in other words, the stochastic ordering of the parameters induces a *partial ordering* over the vector-valued hyperparameters. The transition probabilities $p(\underline{\alpha}_t \mid \underline{\alpha}_{t-1})$ then provide the appropriate measure over $R(\underline{\alpha}_{t-1})$.

We also see, upon reflection, that our formulation can produce many other forms of parameter relationships than (1.2) or (2.1), through appropriate choices of R. For instance, if $p(\theta \mid \underline{\alpha})$ is a

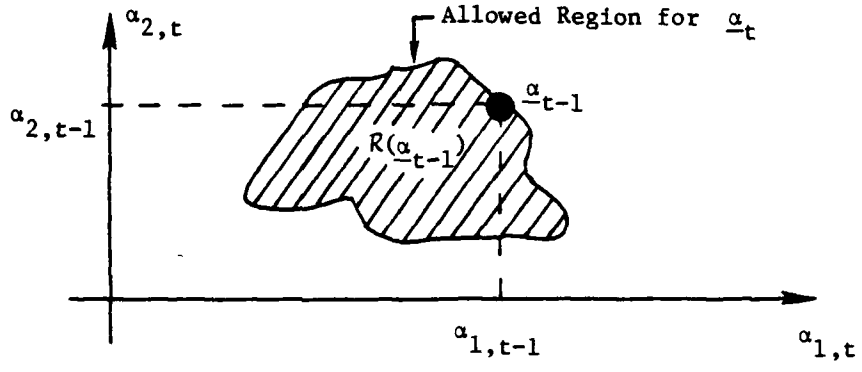


FIGURE 2

MAPPING OF HYPERPARAMETERS FOR STOCHASTIC ORDERING

gamma density, an ordering of the shape parameter, keeping the mean fixed, will give a family F of priors with ordered coefficients of variation, ranging from exponential to degenerate densities. Other variations are left to the reader's imagination.

9. APPLICATION

In spite of its generality, this approach has serious practical drawbacks. The first is that it is difficult to construct analytic models in which the hierarchical computations can be carried out explicitly; even with a natural conjugate prior on θ, there remains the problem of specifying the hyperparameter density in a convenient form. ([4] illustrates some of these problems using the shrinkage (1.4).)

Secondly, digital computations require an additional dimension of the discretized space, so that it is difficult to approximate the usual continuous densities. Limited computational experience with small, rectangular priors and hyperpriors indicated that the growth under stochastic ordering behaves as expected, with an occasional reversal of growth, but otherwise was not illuminating.

Finally, of course, we should question the hypothesis of stochastic ordering itself. A decision-maker might believe that improvement occurs "on the average", for instance, but would prefer to model the anti-growth situations in a completely different way, with, say, (2.1) replaced by an ordering of the means only. We doubt seriously that any real data could discriminate against two such competing models.

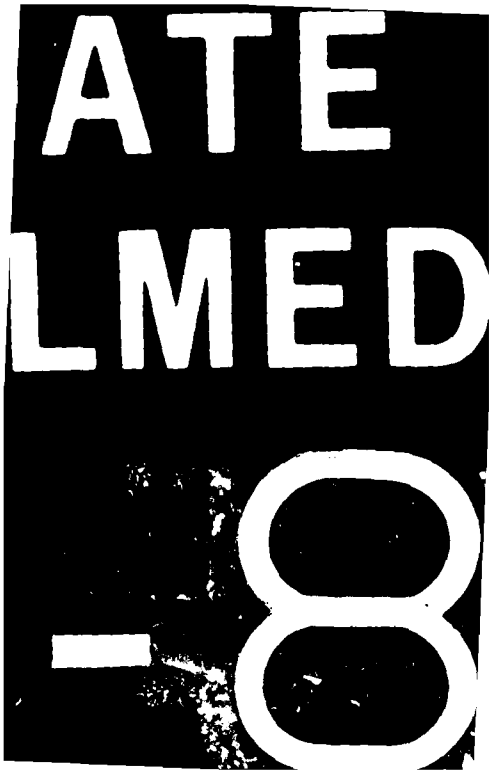
REFERENCES

[1] Jewell, W. S., "Three Papers in Credibility Theory," ORC 76-16, Operations Research Center, University of California, Berkeley, (1976).

[2] Jewell, W. S., "A General Framework for Learning-Curve Reliability Growth Models," ORC 79-11, Operations Research Center, University of California, Berkeley, (1979).

[3] Singpurwalla, N., "An Adaptive Bayesian Scheme for Estimating Reliability Growth Under Exponential Failure Times," ORC 79-17, Operations Research Center, University of California, Berkeley, (1979).

[4] Smith, A. F. M., "A Bayesian Note on Reliability Growth During a Development Testing Program," IEEE Transactions on Reliability, Vol. R-26, pp. 346-347, (1977).


ATE
LMED
-8